Available online at www.sciencedirect.com

SciVerse ScienceDirect

ELSEVIER

Journal of Functional Analysis 263 (2012) 3306–3332

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

A Donsker theorem for Lévy measures

Richard Nickl [a,*], Markus Reiß [b]

[a] Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, CB30WB, Cambridge, UK

[b] Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Received 5 January 2012; accepted 12 August 2012

Available online 12 September 2012

Communicated by B. Driver

Abstract

Given n equidistant realisations of a Lévy process $(L_t,\ t \geqslant 0)$, a natural estimator $\hat{N}_n$ for the distribution function N of the Lévy measure is constructed. Under a polynomial decay restriction on the characteristic function φ, a Donsker-type theorem is proved, that is, a functional central limit theorem for the process $\sqrt{n}(\hat{N}_n - N)$ in the space of bounded functions away from zero. The limit distribution is a generalised Brownian bridge process with bounded and continuous sample paths whose covariance structure depends on the Fourier-integral operator $\mathcal{F}^{-1}[1/\varphi(-\bullet)]$. The class of Lévy processes covered includes several relevant examples such as compound Poisson, Gamma and self-decomposable processes. Main ideas in the proof include establishing pseudo-locality of the Fourier-integral operator and recent techniques from smoothed empirical processes.

Keywords: Uniform central limit theorem; Nonlinear inverse problem; Smoothed empirical processes; Pseudo-differential operators; Jump measure

1. Introduction

A classical result of probability theory is Donsker's central limit theorem for empirical distribution functions: If $X_1, \dots, X_n$ are i.i.d. random variables with distribution function $F(t) = P((-\infty, t]), t \in \mathbb{R}$, and if $F_n(t) = P_n((-\infty, t])$ where $P_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ is the empirical

* Corresponding author.
E-mail addresses: r.nickl@statslab.cam.ac.uk (R. Nickl), mreiss@math.hu-berlin.de (M. Reiß).

0022-1236/$ – see front matter
http://dx.doi.org/10.1016/j.jfa.2012.08.012

measure, then $\sqrt{n}(F_n - F)$ converges in law in the Banach space of bounded functions on $\mathbb{R}$, to a P-Brownian bridge. The result in itself and its many extensions have been at the heart of much of our understanding of modern statistics, see the monographs [8,27] for a comprehensive account of the foundations of this theory.

The purpose of this article is to investigate a conceptually closely related problem: at equidistant time steps $t_k = k\Delta$, $k = 0, 1, \dots, n$, one observes a trajectory of a Lévy process with corresponding Lévy (or jump) measure ν, and wishes to estimate the distribution function N of ν. Since we do not assume that the time distance Δ varies (in particular, no high-frequency regime), we equivalently observe a sample from an infinitely divisible distribution given by the i.i.d. increments of the process. Since ν is only a finite measure away from zero the natural target of estimation is $N(t) = \nu((-\infty, t])$ for $t < 0$ and $N(t) = \nu([t, \infty))$ for $t > 0$. By analogy to the classical case of estimating F, one aims for an estimator $\hat{N}$ such that $\sqrt{n}(\hat{N} - N)$ satisfies a limit theorem in the space of functions bounded on $\mathbb{R} \setminus (-\zeta, \zeta)$, $\zeta > 0$. Statistical minimax theory reveals that the problem of estimating N is intrinsically more difficult than the one of estimating F – it is a nonlinear inverse problem in the terminology of nonparametric statistics. We discuss this point in more detail below, but note that it implies that a rate of convergence $1/\sqrt{n}$ for $\hat{N}(t) - N(t)$, even only at a single point t, cannot be achieved (by any estimator $\hat{N}$) without certain qualitative assumptions on the Lévy process. Particularly, the process cannot contain a nonzero Gaussian component. On the other hand, and perhaps surprisingly, we show in the present article that for a large and relevant class of Lévy processes a Donsker theorem can be proved.

Similar to Donsker's classical theorem our results have interesting consequences for statistical inference, such as the construction of confidence bands and goodness of fit tests. While we do not address these issues explicitly here and concentrate on spelling out the mathematical ideas, it is nevertheless instructive to discuss some related literature on statistical inference on the Lévy triplet from discrete observations. The basic principle for understanding the nonlinearity in this setting is already inherent in the problem of *decompounding* a compound Poisson process, which has been studied in queuing theory and insurance mathematics. In this case the Lévy measure ν is a finite measure and by explicit inversion in the convolution algebra Buchmann and Grübel [4] prove a central limit theorem with rate $1/\sqrt{n}$ for a plug-in estimator of N in an exponentially weighted supremum norm, assuming that the intensity of the process is known.

For general Lévy triplets the estimation problem is generally ill-posed in the sense of inverse problems. In fact, the linearised problem is of deconvolution-type where the part of the error distribution is taken over by the observation law itself, see Eq. (4.1) below. This phenomenon, which could be coined *auto-deconvolution*, was first studied by Belomestny and Reiß [3]. For the general problem of estimating functionals of the Lévy measure the results by Neumann and Reiß [19] show in particular that a functional can be estimated at parametric rate $1/\sqrt{n}$ provided its smoothness outweighs the ill-posedness induced by the decay of the characteristic function. Comparing to [19] we are thus interested in the low regularity functional $f \mapsto \int_{-\infty}^{t} f$ (not covered by their results), and in exact limiting distributions. Instead of making inference on the distribution function, one may also be interested in the associated nonparametric estimation problem for a Lebesgue density of the Lévy measure, where the rate $1/\sqrt{n}$ can never be attained. This problem was studied in [14] for Lévy processes with finite jump activity and a Gaussian part, [5] for a model selection procedure in the finite variation case, or [25] for self-decomposable processes. Generalisations for observations of more general jump processes like Lévy–Ornstein–Uhlenbeck processes or affine processes are considered by Jongbloed, van der Meulen and van der Vaart [15] and Belomestny [2].

The proof of our main result contains certain subtleties that we wish to briefly discuss here: In the classical Donsker case one proves that the empirical process $\sqrt{n}(P_n - P)$ is tight in the space of bounded mappings acting on $\{1_{(-\infty,t]}\colon t \in \mathbb{R}\}$. The ill-posedness of the Lévy-problem can be roughly understood, after linearisation, as requiring to show that the empirical process $\sqrt{n}(P_n - P)$ is tight in the space of bounded mappings acting on the class

$$\mathcal{G}_\varphi = \bigl\{\mathcal{F}^{-1}\bigl[1/\varphi(-\bullet)\bigr] * 1_{(-\infty,t]}\colon |t| \geqslant \zeta\bigr\}, \tag{1.1}$$

where $\zeta > 0$ is arbitrary, $\mathcal{F}$ is the Fourier transform and where $\varphi = \mathcal{F}P$ is the characteristic function of the increments of the Lévy process. In fact, the situation is more complicated than that, but the above simplification highlights the main problem. Convolution with $\mathcal{F}^{-1}[1/\varphi]$ is just a way of writing deconvolution with $P = \mathcal{F}^{-1}[\varphi]$, which is mathematically understood as the action of a pseudo-differential operator, and the class $\mathcal{G}_\varphi$ can be shown *not* to be P-Donsker (arguing as in Theorem 7 in [20], for instance), unless in very specific situations (effectively in the compound Poisson case discussed above). In other words, the empirical process is not tight when indexed by these functions.

A starting point of our analysis is that for certain Lévy processes a generalised P-Brownian bridge $\mathbb{G}^\varphi$ with bounded sample paths can be defined on $\mathcal{G}_\varphi$, uniformly continuous for the intrinsic covariance metric of $\mathbb{G}^\varphi$, see Theorem 9. Roughly speaking this means that a tight limit process exists, and that a limit theorem at rate $1/\sqrt{n}$ may hold if one replaces the empirical process by a smoothed one. This hope is nourished by the phenomenon – first observed, in a general empirical process setting unrelated to the present situation, by Radulović and Wegkamp [22], and recently developed further in several directions by Giné and Nickl [10] – that smoothed empirical processes may converge in situations where the unsmoothed process does not. The results in [10] apply to unbounded classes, so in particular to $\mathcal{G}_\varphi$, and this idea in combination with a thorough analysis of the pseudo-differential operator $\mathcal{F}^{-1}[1/\varphi(-\bullet)]$ are at the heart of our proofs.

The paper is organised as follows: Section 2 contains the exact conditions on the model, the construction of the estimator and the main result. In Section 3 the model assumptions, some important examples and potential extensions are discussed. Finally, the complete proof of the Donsker-type result is given in Section 4, divided into the finite-dimensional central limit theorem and the uniform tightness result.

2. The setting and main result

We observe a real-valued Lévy process $(L_t,\ t \geqslant 0)$ at equidistant time points $t_k = k\Delta$, $k = 0, 1, \ldots, n$, for $\Delta > 0$ fixed. It will be seen to be natural (Section 3) to restrict to Lévy processes of (locally) finite variation. In this case the characteristic function of the increments $X_k := L_{t_k} - L_{t_{k-1}}$ is given by

$$\varphi(u) = \mathbb{E}\bigl[\exp(iuL_\Delta)\bigr] = e^{\Delta\psi(u)} \quad \text{where } \psi(u) = i\gamma u + \int\limits_{\mathbb{R}\setminus\{0\}} \bigl(e^{iux} - 1\bigr)\,\nu(dx)$$

with drift parameter $\gamma \in \mathbb{R}$ and Lévy (or jump) measure ν satisfying $\int_{\mathbb{R}}(|x| \wedge 1)\,\nu(dx) < \infty$ (due to finite variation). The increments $X_1, \ldots, X_n$ are i.i.d. and we write P for the law of X_k and p for its density (if it exists) as well as $P_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ and $\varphi_n(u) = \mathcal{F}P_n(u) = \int e^{iux}\,dP_n(x)$ for the empirical measure and empirical characteristic function, respectively. Throughout $\mathcal{F}$ denotes the Fourier(–Plancherel) transform acting on finite measures, on the space $L^1(\mathbb{R})$ of integrable or on the space $L^2(\mathbb{R})$ of square-integrable functions on $\mathbb{R}$, see e.g. [16] for the standard Fourier techniques that we shall employ.

If ν has a finite first moment, then the weighted Lévy measure $x\,\nu(dx)$ can be identified directly from the law of X_k in the Fourier domain:

$$\frac{1}{i\Delta}\frac{\varphi'(u)}{\varphi(u)} = -i\psi'(u) = \gamma + \int e^{iux} x\,\nu(dx) = \gamma + \mathcal{F}[x\nu](u). \tag{2.1}$$

Our goal is to estimate the cumulative distribution function of ν,

$$N(t) := \begin{cases} \nu((-\infty,t]), & t<0,\\ \nu([t,\infty)), & t>0, \end{cases} \tag{2.2}$$

from the sample $X_1,\dots,X_n$. Note that in general $N(t)$ tends to infinity for $t\to 0$. If we denote by $\mathcal{F}^{-1}$ the inverse Fourier transform, then the relation (2.1) suggests a natural empirical estimate of $N(t)$ (we shall see below that γ can be neglected),

$$\hat{N}_n(t) := \int_{\mathbb{R}} g_t(x)\mathcal{F}^{-1}\left[\frac{1}{i\Delta}\frac{\varphi_n'}{\varphi_n}\mathcal{F}K_h\right](x)\,dx$$
$$\text{with } g_t(x) := \begin{cases} x^{-1}1_{(-\infty,t]}(x), & t<0,\\ x^{-1}1_{[t,\infty)}(x), & t>0, \end{cases} \tag{2.3}$$

where K is a band-limited kernel function and $K_h(x) := h^{-1}K(x/h)$. In the sequel the kernel will be required to satisfy

$$\int K = 1, \qquad \operatorname{supp}(\mathcal{F}K)\subseteq[-1,1] \quad \text{and}$$
$$|K(x)| + |K'(x)| \lesssim (1+|x|)^{-\beta} \quad \text{for some } \beta > 2. \tag{2.4}$$

Throughout, we shall write $A_p \lesssim B_p$ if $A_p \leqslant CB_p$ holds with a uniform constant C in the parameter p as well as $A_p \sim B_p$ if $A_p \lesssim B_p$ and $B_p \lesssim A_p$.

The smooth spectral cutoff induced by multiplication with $\mathcal{F}K_h$ is desirable for various reasons; in particular, it will imply that $\hat{N}_n$ is well-defined with probability tending to one. By Plancherel's formula, we have the alternative representation

$$\hat{N}_n(t) := \frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g_t(-u)\frac{\varphi_n'(u)}{\varphi_n(u)}\mathcal{F}K_h(u)\,du.$$

Heuristically, for $h_n\to 0$ we expect consistency $\hat{N}_n(t)\to N(t)$ in probability, $t\neq 0$, because as $h_n\to 0$ we have $K_{h_n}\to\delta_0$ (the Dirac measure in zero) and thus $\mathcal{F}K_{h_n}(u)\to 1$ which may be combined with the law of large numbers for both φ_n and φ_n'. For this argument to work it is important to note that the drift γ induces a point measure in zero for $\mathcal{F}^{-1}[\varphi'/\varphi]$ which is outside the support of g_t, cf. Section 4.1.1 below. For our precise results we shall need the following conditions on the data-generating Lévy process. Throughout the paper we often write φ^{-1} for $1/\varphi$.

Assumption 1. We require for some $\varepsilon > 0$:

(a) $\int \max(|x|, |x|^{2+\varepsilon})\, \nu(dx) < \infty$;
(b) $x\nu$ has a bounded Lebesgue density and $|\mathcal{F}[x\nu](u)| \lesssim (1+|u|)^{-1}$;
(c) $(1+|u|)^{-1+\varepsilon}\varphi^{-1}(u) \in L^2(\mathbb{R})$.

Assumption 1(a) imposes finite variation, ensuring the identification identity (2.1), as well as finite $(2+\varepsilon)$-moments of ν and P, since by Theorem 25.3 in [23]

$$\int_{\mathbb{R}} |x|^{2+\varepsilon}\, \nu(dx) < \infty \quad \Leftrightarrow \quad \int_{\mathbb{R}} |x|^{2+\varepsilon}\, P(dx) < \infty. \tag{2.5}$$

As $\hat{N}$ is based on $\varphi_n'(u)$, and since a central limit theorem is desired, it is natural to require a finite second moment of X_k. The additional ε in the power will allow to apply the Lyapounov criterion in the CLT for triangular schemes and to obtain uniform in u stochastic bounds for $\varphi_n'(u) - \varphi'(u)$ over increasing intervals. Assumptions 1(b) and 1(c) are discussed in more detail after the following theorem, which is the main result of this article.

For $\zeta > 0$, let $\ell^\infty((-\infty, -\zeta] \cup [\zeta, \infty))$ be the space of bounded real-valued functions on $(-\infty, -\zeta] \cup [\zeta, \infty)$ equipped with the supremum norm. Convergence in law in this space, denoted by $\to^{\mathcal{L}}$, is defined as in [8, p. 94].

Theorem 2. *Suppose that Assumption 1 is satisfied, $\zeta > 0$ and $h_n \sim n^{-1/2}(\log n)^{-\rho}$ for some $\rho > 1$. Then as $n \to \infty$*

$$\sqrt{n}(\hat{N}_n - N) \to^{\mathcal{L}} \mathbb{G}^\varphi \quad \text{in } \ell^\infty\big((-\infty, -\zeta] \cup [\zeta, \infty)\big),$$

where $\mathbb{G}^\varphi$ is a centered Gaussian Borel random variable in $\ell^\infty((-\infty, -\zeta] \cup [\zeta, \infty))$ with covariance structure given by

$$\Sigma_{t,s} = \frac{1}{\Delta^2} \int_{\mathbb{R}} \left(\mathcal{F}^{-1}\left[\frac{1}{\varphi(-\bullet)}\right] * \big(x g_t(x)\big)\right) \times \left(\mathcal{F}^{-1}\left[\frac{1}{\varphi(-\bullet)}\right] * \big(x g_s(x)\big)\right) P(dx)$$

and where g_t is given in (2.3).

In view of $xg_t(x) = 1_{(-\infty,t]}(x)$ for $t < 0$ and symmetrically for $t > 0$, the representation of the covariance in the theorem above is intuitively appealing when compared to the classical Donsker theorem. Its rigorous interpretation, however, needs some care, as it is not quite clear how the pseudo-differential operator $\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)]$ acts on the indicator function $xg_t(x)$. One rigorous representation that follows from our proofs uses

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-\bullet)\big] * 1_{(-\infty,t]} = \mathcal{F}^{-1}\big[(1+iu)^{-1}\varphi^{-1}(-u)\big] * (1_{(-\infty,t]} + \delta_t)$$

together with the fact that $\mathcal{F}^{-1}[(1+iu)^{-1}\varphi^{-1}(-u)]$ can be shown to be contained in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ under Assumption 1 (using lifting properties of Besov spaces), so that the right-hand side of the last display is defined almost everywhere.

Another more explicit representation, which also implies that $\Sigma_{t,t} < \infty$, is the following: Note that formally

$$\int_{\mathbb{R}} \mathcal{F}^{-1}\left[\frac{1}{\varphi(-\bullet)}\right] * \big(xg_t(x)\big)\,dP(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \big(\mathcal{F}[xg_t](-\bullet)\big)(u)\varphi^{-1}(u)\varphi(u)\,du = (xg_t)(0) = 0,$$

which explains why the covariance in Theorem 2 is centered for $t \neq 0$. Moreover, $\mathcal{F}[xg_t] = i^{-1}(\mathcal{F}[g_t])'$ and integration by parts gives rise to the formally equivalent representation

$$\Sigma_{t,s} = (i\Delta)^{-2}\int_{\mathbb{R}} h_t(x)h_s(x)\,P(dx) \tag{2.6}$$

where

$$h_t(x) = \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t(u)\big](x)ix + \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g_t(u)\big](x),$$

and where we note that $i^{-1}h_t$ is real-valued. This expression for h_t is the one we shall employ in our proofs, as it can be shown to be rigorously defined in $L^2(P)$ under the maintained assumptions, see (4.11) below for more details.

Moreover the last representation immediately suggests consistent estimators of $\Sigma_{t,s}$ based on the empirical characteristic function φ_n and the empirical measure P_n, useful when one is interested in the Gaussian limiting distribution for inference purposes on N.

3. Discussion

3.1. The regularity conditions

We remark first that the results in [19] imply that we can attain a $1/\sqrt{n}$-rate for estimation only if the characteristic function decays at most with a low polynomial order. This restricts the classes of Lévy processes automatically to the (locally) finite variation case (e.g. proof of Proposition 28.3 in [23]), and moreover excludes all Lévy processes with a nonzero Gaussian component.

Let us next discuss Assumption 1(c) which describes the lower bound we need on the ill-posedness of the estimation problem. It holds for all compound Poisson processes, in which case $|\varphi^{-1}(u)|$ is bounded, but also for Gamma processes with $\alpha \in (0, 1/(2\Delta))$ and for pure-jump self-decomposable processes with not too high jump activity at zero, see Proposition 3 below. Recall (e.g. [23, Section 15]) that self-decomposable distributions describe the limit laws of suitably rescaled sums of independent random variables as well as the stationary distributions of Lévy–Ornstein–Uhlenbeck processes, and thus give rise to a rich nonparametric class of Lévy measures. More generally, if $\mathbb{E}[e^{iuL_1}]$ decays polynomially, then there exists a $\Delta_0 > 0$ such that for all $\Delta < \Delta_0$ the corresponding characteristic function $\varphi(u) = \mathbb{E}[e^{iuL_\Delta}]$ satisfies $|\varphi^{-1}(u)| \lesssim (1+|u|)^\alpha$ for $\alpha < 1/2$, so Assumption 1(c) holds for any polynomially decaying φ if the sampling frequency is large (i.e., Δ small) enough. Abstractly, Assumption 1(c) means that the pseudo-differential operator $\mathcal{F}^{-1}[\varphi^{-1}]$ of deconvolution is an element of the L^2-Sobolev space $H^{-1+\varepsilon}(\mathbb{R})$ of negative order $\varepsilon - 1$. In the simpler problem of statistical deconvolution an analogous restriction for the characteristic function of the error variables is necessary, even if one is only interested in rates of convergence of an estimator, and the situation is similar here: The lower bound techniques from Theorem 4.4 of [19] or Theorem 1 of [18] can be adapted to the present situation to imply, for instance, that for Gamma processes with $\alpha > 1/(2\Delta)$ the 'parametric' rate $1/\sqrt{n}$ cannot be achieved by any estimator in the Lévy estimation problem considered here, so that Assumption 1(c) is in this sense sharp for Theorem 2.

The smoothness condition on $x\nu$ in Assumption 1(b) is not very restrictive: it is satisfied whenever the weighted Lévy measure $x\nu$ has a density whose weak derivative is a finite measure (noting $x\nu \in L^1(\mathbb{R})$ by Assumption 1(a)). As simple examples, any compound Poisson process with a jump density of bounded variation and a finite first moment satisfies this condition, as does any Gamma process. More generally, most self-decomposable processes satisfy this condition, see Proposition 3 below.

The key role of Assumption 1(b) is not to enforce smoothness of ν, but to ensure *pseudo-locality* of the deconvolution operator $\mathcal{F}^{-1}[\varphi^{-1}]$ in the sense that the location of singularities like the jump in the indicator $1_{(-\infty,t]}$ remains unchanged under deconvolution. A similar situation arises in standard deconvolution problems, see the recent paper [24]. In the spirit of the theory of pseudo-differential operators this is established by differentiating in the spectral domain, see (4.10) below for details,

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] = \frac{1}{i\bullet}\mathcal{F}^{-1}\big[\big(\varphi^{-1}(-u)\big)'\big]$$

under the condition that $(\varphi^{-1})' = \Delta\psi'\varphi^{-1} \in L^2(\mathbb{R})$. Neglecting the drift, ψ' is $\mathcal{F}[ix\nu]$ and Assumptions 1(b), 1(c) together ensure $(\varphi^{-1})' \in L^2(\mathbb{R})$, see Lemma 4 below. As discussed later, the example of a superposition of a Gamma and a Poisson process provides a simple concrete situation where a violation of this condition renders the asymptotic variance in Theorem 2 infinite.

There is another interesting interaction between Assumptions 1(b) and 1(c). A decay rate $|u|^{-1}$ for $\mathcal{F}[x\nu](u)$ is the maximal possible smoothness requirement under Assumption 1(c); otherwise $|\mathrm{Re}(\psi'(u))| \leqslant |\mathcal{F}[x\nu](u)| = o(|u|^{-1})$ would imply $|\varphi(u)| = \exp(\mathrm{Re}(\Delta\psi(u))) = \exp(o(\log(u)))$ for $|u| \to \infty$, excluding polynomial decay of the characteristic function φ.

3.2. Examples

We now discuss a few examples in more detail.

Compound Poisson processes. The compound Poisson case where ν is a finite measure is covered in Theorem 2. Note that due to the presence of a point mass at zero in P the characteristic function satisfies $\inf_u |\varphi(u)| \geqslant \exp(-2\nu(\mathbb{R})) > 0$ $(\Delta = 1)$. Therefore Assumption 1(c) is trivially satisfied. Assumption 1(b) requires that the law of the jump sizes has a density ν such that $x\nu(x)$ is bounded and has the respective decay property in the Fourier domain. Assumption 1(a) just postulates $(2+\varepsilon)$ finite moments of the jump law. Compared to [4] we thus obtain directly a uniform central limit without weighting, exponential moments and, perhaps more importantly, without prior knowledge of the intensity, yet our result holds only away from the origin and under Assumption 1(b).

Stronger results can be obtained by adapting our method to this specific case because the distribution function N of ν is defined classically for all $t \in \mathbb{R}$ and Assumption 1(b) is not required to ensure pseudo-locality of deconvolution. In fact, deconvolution reduces to convolution with a signed measure because of ($\bar{\nu}^{*k}$ denotes k-fold convolution)

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-\bullet)\big] = \sum_{k=0}^{\infty} \frac{e^{\lambda}(-1)^k}{k!}\bar{\nu}^{*k} \quad \text{with } \lambda := \nu(\mathbb{R}),\ \bar{\nu}(A) := \nu(-A).$$

Therefore, $\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)] * 1_{(-\infty,t]}$ is a bounded function, in fact of bounded variation, and the uniform CLT for the linearised stochastic term follows directly (since BV-balls are universal Donsker classes). The remainder term remains negligible whenever the inverse bandwidth h^{-1} grows slower than exponentially in n. Choosing for instance $h_n \sim \exp(-\sqrt{n})$ yields a pointwise CLT for $\sqrt{n}(\hat{N}_n(t) - N(t))$ for all $t \in \mathbb{R}$ if the bias is negligible, e.g. if N has some positive Hölder regularity at t. We do not pursue a detailed derivation of this specific case here.

Gamma processes. The family of Gamma processes satisfies $X_k \sim \Gamma(\alpha\Delta, \lambda)$, with probability density $\gamma(y; \alpha\Delta, \lambda) = (1/\Gamma(\alpha\Delta))\lambda^{\alpha\Delta} y^{\alpha\Delta-1} e^{-\lambda y}$, Lévy measure $\nu(dx) = \alpha x^{-1} e^{-\lambda x} 1_{\mathbb{R}^+}(x)\,dx$ and characteristic function $\varphi(u) = (1 - iu/\lambda)^{-\alpha\Delta} 1_{\mathbb{R}^+}(y)$. For simplicity we consider $\lambda = 1$ and, in order to satisfy Assumption 1(c), we restrict to $\alpha \in (0, 1/(2\Delta))$. We denote the density of $\Gamma(\beta, 1)$ by γ_β and its distribution function by Γ_β. Then

$$\mathcal{F}^{-1}\big[\varphi^{-1}\big] = \mathcal{F}^{-1}\big[(1 - iu)^{\alpha\Delta-1}(1 - iu)\big] = \gamma_{1-\alpha\Delta} * (\mathrm{Id} + D)$$

holds with the differential operator D. This is a well-known form of the fractional derivative operator of order $\alpha\Delta$. We deduce

$$\mathcal{F}^{-1}\big[\varphi^{-1}(-\bullet)\big] * 1_{[t,\infty)} = \gamma_{1-\alpha\Delta}(-\bullet) * (1_{[t,\infty)} - \delta_t).$$

Hence, for $t > 0$ the asymptotic variance of Theorem 2 is given by

$$\Sigma_{t,t} = \int_0^\infty \big(1 - \Gamma_{1-\alpha\Delta}(t - x) - \gamma_{1-\alpha\Delta}(t - x)\big)^2 \gamma_{\alpha\Delta}(x)\,dx.$$

Note that the integrand has poles of order $(\alpha\Delta)^2$ at $x = t$ and of order $1 - \alpha\Delta$ at $x = 0$ such that the variance is finite if and only if $\alpha\Delta < 1/2$ and $t \neq 0$. So, in this case, Assumption 1(c) prevents Σ_{tt} from being infinite.

The Gamma process case can serve as a basic example for all the theory that follows. It reveals the problem that standard L^p-theory or non-local Fourier analysis will not be sufficient in this context as different locations of the singular support (the poles) are required to ensure finiteness of $\Sigma_{t,t}$.

Gamma plus Poisson process. Let us briefly give a simple counterexample showing that pseudo-locality of the deconvolution operator is important. If the Lévy process is a superposition of a Gamma process as above with $\alpha \in (0, 1/(2\Delta))$ and of an independent Poisson process with intensity $\lambda > 0$, the density p of the increments is given by the convolution of the $\gamma_{\alpha\Delta}$-density with a Poiss(λ)-law and thus has poles of order $1 - \alpha\Delta$ at $x \in \mathbb{N}_0$. On the other hand, the deconvolution operator is given by

$$\mathcal{F}^{-1}\left[\varphi^{-1}(-\bullet)\right] = \sum_{k=0}^{\infty} \frac{e^{\lambda}(-1)^k}{k!} \delta_{-k} * \gamma_{1-\alpha\Delta}(-\bullet) * (\mathrm{Id} - D)$$
$$= \sum_{k=0}^{\infty} \frac{e^{\lambda}(-1)^k}{k!} \gamma_{1-\alpha\Delta}(-\bullet - k) * (\mathrm{Id} - D).$$

As in the pure Gamma case, this shows that $\Sigma_{t,t}$ is finite if and only if none of the poles at $x = t - k$, $k \in \mathbb{N}_0$, and at $x = k$, $k \in \mathbb{N}_0$, of the respective functions coincide, which is the case only for non-integer $t \notin \mathbb{N}_0$. Consequently, we cannot hope even to prove a pointwise CLT with rate $1/\sqrt{n}$ at integers t. This case that singularities are just translated by convolution with point measures is excluded by the regularity requirement for $x\nu$ in Assumption 1(b).

Self-decomposable processes. We finally consider the class of self-decomposable processes, cf. [23], Section 15, which contains all Gamma processes. For any pure-jump self-decomposable process we have $\nu(dx) = k(x)/|x|\,dx$ with a unimodal k-function increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$. If the limits $k(0-)$ and $k(0+)$ of k at zero are finite, then k is a function of bounded variation and so is $\operatorname{sgn}(x)k(x)$, the density of $x\nu$. The moment condition of Assumption 1(a) in particular implies $\operatorname{sgn}(x)k(x) \in L^1(\mathbb{R})$ which yields Assumption 1(b). It is quite remarkable that the probabilistic property of self-decomposability implies the analytic property of pseudo-locality for the deconvolution operator.

For the characteristic function of self-decomposable processes we have $|\varphi(u)| \gtrsim (1 + |u|)^{-\alpha\Delta}$ with $\alpha = k(0-) + k(0+)$, which follows exactly as the proof of Lemma 2.1 in [25]. The latter is the counterpart to Lemma 53.9 in [23], where an upper bound of the same order times a logarithmic factor is shown. We conclude that Assumption 1(c) translates to the condition $\alpha < 1/(2\Delta)$.

We note that Assumptions 1(a) and 1(b) remain true under superposition of independent Lévy processes and we collect the findings in an explicit statement.

Proposition 3. *Assumption 1 is satisfied for*

(a) *a compound Poisson process whenever the jump law has a density ν such that $x\nu$ is of bounded variation and ν has a finite $(2 + \varepsilon)$-moment,*
(b) *a Gamma process with parameters $\alpha \in (0, 1/(2\Delta))$ and $\lambda > 0$,*
(c) *a pure-jump self-decomposable process whenever its k-function satisfies $\int \max(1, |x|^{1+\varepsilon}) \times k(x)\,dx < \infty$ and $k(0-) + k(0+) < 1/(2\Delta)$,*
(d) *and for any Lévy process which is a sum of independent compound Poisson and self-decomposable processes of the preceding types.*

3.3. Extensions and perspectives

There are many directions for further investigation. As from the classical Donsker result, concrete statistical inference procedures, like Lévy-analogues of the classical Kolmogorov–Smirnov-tests and corresponding confidence bands, can be derived from Theorem 2. Also extensions to uniform CLTs for more general functionals than just for the distribution function are highly relevant. A question of particular interest in the area of statistics for stochastic processes is whether one can allow for high-frequency observation regimes $\Delta_n \to 0$. As discussed above, decreasing $\Delta \to 0$ renders the inverse problem more regular, as Assumption 1(c) is then easier to satisfy. Since we use the central limit theorem for triangular arrays in our proofs, allowing Δ to depend on n should not pose a principal difficulty, but doing so in a sharp way may not only require an estimator based on the second derivative of $\log(\varphi_n)$, but also extra care in controlling all terms uniformly in n, and is beyond the scope of the present paper.

Another issue of statistical relevance is the question of efficiency, which we briefly address here. Our plug-in estimation method is quite natural and should have asymptotic optimality properties as the empirical distribution function has for the classical i.i.d. case. This is also in line with the result by Klaassen and Veerman [17] who show that the tangent space of the class of infinitely divisible distributions with positive Gaussian part is nonparametric to the effect that the estimation of linear functionals $\int g\,dP$ of P (but not ν as in our case) by empirical means is asymptotically efficient. Indeed, a formal derivation indicates that the pointwise asymptotic variance of our estimator $\hat{N}_n(t)$ coincides with the semi-parametric Cramér–Rao information bound (see [27, Chapter 3.11], for the relevant definitions). Let us restrict here to the case $t < 0$ and assume that the observation law P_ν has a Lebesgue density p_ν.

Perturbing the Lévy measure ν in direction of an L^1-function h, we obtain by differentiating in the Fourier domain the score function (the derivative of the log-likelihood)

$$\dot{\ell}_\nu(h) := \frac{d}{d\varepsilon}\frac{p_{\nu+\varepsilon h}}{p_\nu}\bigg|_{\varepsilon=0} = \frac{\mathcal{F}^{-1}[\varphi_\nu(u)\int(e^{iux}-1)\,h(dx)]}{p_\nu} = \frac{p_\nu * (h-\lambda_h\delta_0)}{p_\nu}$$

with $\lambda_h = \int h$. This yields the Fisher information at measure ν in direction h as

$$\big\langle I(\nu)h,h\big\rangle := \mathbb{E}_\nu\big[\dot{\ell}_\nu(h)^2\big] = \int\left(\frac{p_\nu * (h-\lambda_h\delta_0)(x)}{p_\nu(x)}\right)^2 P_\nu(dx).$$

On the other hand, we aim at estimation of the functional $\nu \mapsto N(t)$ whose derivative in direction h by linearity is given by $H(t) = \langle 1_{(-\infty,t]}, h\rangle$ (interpreting $\langle\bullet,\bullet\rangle$ as a dual pairing). The semi-parametric Cramér–Rao lower bound is then $\sup_h \frac{H(t)^2}{\langle I(\nu)h,h\rangle}$, maximising the parametric bound over all sub-models $(\nu+\varepsilon h)_{\varepsilon\in\mathbb{R}}$. The supremum is formally attained at $h^* = I(\nu)^{-1}1_{(-\infty,t]}$ with value $\langle 1_{(-\infty,t]}, h^*\rangle$. The maximiser can be expressed explicitly using the deconvolution operator:

$$h^* = \mathcal{F}^{-1}\big[\varphi^{-1}\big] * \big\{p_\nu \times \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * 1_{(-\infty,t]} - \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * 1_{(-\infty,t]}(0)\big)\big\}.$$

Resuming the formal calculus and noting that $\mathcal{F}^{-1}[\varphi^{-1}(-u)]$ is the formal adjoint of $\mathcal{F}^{-1}[\varphi^{-1}]$, we find the explicit Cramér–Rao bound

$$\begin{aligned}&\int 1_{(-\infty,t]}(x)h^*(x)\,dx\\ &\quad= \int \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * 1_{(-\infty,t]}\big)(x)\,p_\nu(x)\big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * 1_{(-\infty,t]}\big)(x)\,dx,\end{aligned}$$

which is exactly equal to the asymptotic variance $\Sigma_{t,t}$ from Theorem 2. We have used here that $\mathcal{F}^{-1}[\varphi^{-1}(-u)] * 1_{(-\infty,t]}(X)$ is centred, cf. (4.3) below.

The hardest parametric subproblem of our general semi-parametric estimation problem is thus given by perturbing ν in direction of h^*. The lower bound for the variance equals exactly the asymptotic variance of our estimator. Let us nevertheless emphasise that this formal derivation of the Cramér–Rao lower bound does not justify asymptotic efficiency in a completely rigorous manner: for this one would have to establish the regularity of the statistical model and $h^* \in L^1(\mathbb{R})$, which appears to require an even finer analysis of the main terms than our Donsker-type result. The complete proof remains a challenging open problem.

4. Proof of Theorem 2

The remainder of this article is devoted to the proof of Theorem 2, which is split into the separate proofs of convergence of the finite-dimensional distributions and of tightness. We shall repeatedly use the following auxiliary lemma.

Lemma 4. *Suppose $\gamma = 0$. Then Assumption 1 implies*:

(a) *The measure $xP = x\,P(dx)$ has a bounded Lebesgue density on $\mathbb{R}$.*
(b) $(\varphi^{-1})' \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R})$ *as well as* $|\varphi^{-1}(u)| \lesssim (1+|u|)^{(1-\varepsilon)/2}$ *for all* $u \in \mathbb{R}$.
(c) $m(u) := \varphi^{-1}(-u)(1+iu)^{(-1+\varepsilon)/2}$ *is a Fourier multiplier on every Besov space* $B^s_{p,q}(\mathbb{R})$ *with* $s \in \mathbb{R}$, $p, q \in [1, \infty]$; *that is convolution with* $\mathcal{F}^{-1}m$ *is continuous from* $B^s_{p,q}(\mathbb{R})$ *to* $B^s_{p,q}(\mathbb{R})$.

Proof. (a) From (2.1) with $\gamma = 0$ we see

$$\mathcal{F}[ixP](u) = \varphi'(u) = i\Delta\mathcal{F}[x\nu](u)\mathcal{F}P(u) \quad \Rightarrow \quad xP = \Delta(x\nu) * P \tag{4.1}$$

and thus with $x\nu$ (Assumption 1(b)) also xP has a Lebesgue density $xp(x)$ with $\|xp\|_\infty \leqslant \Delta\|x\nu\|_\infty$.

(b) From Assumption 1(b) and $\gamma = 0$ we deduce $|\psi'(u)| \lesssim (1+|u|)^{-1}$ and thus $\|(1+|u|)^\varepsilon(\varphi^{-1})'\|_{L^2} \lesssim \|\varphi^{-1}(1+|u|)^{-1+\varepsilon}\|_{L^2} < \infty$ by Assumption 1(c). This implies

$$\begin{aligned}
\left|\varphi^{-1}(u)\right| &\leqslant 1 + \int_0^u \left|(\varphi^{-1})'(v)\right| dv \lesssim 1 + \left\|\left(1+|v|\right)^\varepsilon\left(\varphi^{-1}\right)'\right\|_{L^2}\left\|\left(1+|v|\right)^{-\varepsilon}1_{[0,u]}\right\|_{L^2} \\
&\lesssim \left(1+|u|\right)^{(1/2)-\varepsilon} \lesssim \left(1+|u|\right)^{(1-\varepsilon)/2},
\end{aligned}$$

and then also $|(\varphi^{-1})'|(u) \lesssim |\varphi^{-1}(u)||\psi'(u)| \lesssim 1$, so $(\varphi^{-1})' \in L^\infty(\mathbb{R})$.

(c) The Fourier multiplier property of m follows from the Mihlin multiplier theorem for Besov spaces (see e.g. [26] and particularly the scalar version of Corollary 4.11(b) in [13]): because of (b) the function m is bounded and satisfies

$$\left|um'(u)\right| \lesssim \left|um(u)\right|\left(1+|u|\right)^{-1} \lesssim 1.$$

Consequently, the conditions of Mihlin's multiplier theorem are fulfilled and m is a Fourier multiplier on all Besov spaces $B^s_{p,q}(\mathbb{R})$. $\square$

4.1. Convergence of the finite-dimensional distributions

Denote by $H^s(\mathbb{R})$, $s \in \mathbb{R}$, the standard L^2-Sobolev spaces with norm $\|h\|_{H^s} := \|\mathcal{F}h(u)(1+|u|)^s\|_{L^2}$.

Definition 5. We say that a function $g \in L^\infty(\mathbb{R}) \cap L^2(\mathbb{R})$ is *admissible* if

(a) g is Lipschitz continuous in a neighbourhood of zero,
(b) we can split $g = g^c + g^s$ into functions $g^c \in H^1(\mathbb{R})$, $g^s \in L^1(\mathbb{R})$, satisfying $\max(|\mathcal{F}[g^s](u)|, |\mathcal{F}[xg^s](u)|) \lesssim (1+|u|)^{-1}$ for all $u \in \mathbb{R}$.

Lemma 6. *The functions g_t from (2.3) as well as all finite linear combinations $\sum_i \alpha_i g_{t_i}$ with $\alpha_i \in \mathbb{R}$, $t_i \neq 0$, are admissible. Moreover, we can choose g_t^c, g_t^s in such a way that*

$$\|g_t^c\|_{H^1} \lesssim \big(1+|t|\big)^{-1/2}, \qquad \big|\mathcal{F}g_t^s(u)\big| \lesssim \big(1+|u|\big)^{-1}\big(1+|t|\big)^{-1} \quad \textit{and}$$
$$\big|\mathcal{F}\big[xg_t^s\big](u)\big| \lesssim \big(1+|u|\big)^{-1},$$

the inequalities holding with constants independent of $u \in \mathbb{R}$, $t \in \mathbb{R} \setminus (-\zeta, \zeta)$ for $\zeta > 0$ fixed.

Proof. First note that all properties of admissible functions remain invariant under finite linear combinations and reflection $g \mapsto g(-\bullet)$. It thus suffices to check that g_t, $t < 0$, is admissible. Let $\chi \in C^\infty((-\infty, 0])$ be a smooth function with $\chi(0) = 1$ and χ, χ' both bounded and integrable on $(-\infty, 0]$, for instance $\chi(x) = e^x 1_{(-\infty,0]}(x)$. Decompose $g_t = g_t^c + g_t^s$ with

$$g_t^c(x) = g_t(x)\big(1 - \chi(x-t)\big), \qquad g_t^s(x) = g_t(x)\chi(x-t); \quad \text{for } x \leqslant t,$$

and both equal to zero for $x > t$. Then $g_t^c \in L^2(\mathbb{R})$ and its (weak) derivative is

$$\big(g_t^c\big)'(x) = -x^{-2}\big(1 - \chi(x-t)\big)1_{(-\infty,t]}(x) + x^{-1}\big(1 - \chi(x-t)\big)'1_{(-\infty,t]}(x) \in L^2(\mathbb{R}),$$

so $g_t^c \in H^1(\mathbb{R})$. The functions g_t^s, xg_t^s are both integrable since χ is. The (weak) derivatives of xg_t^s and g_t^s are $\chi'(x-t)1_{(-\infty,t)} - \delta_t$ and $-x^{-2}\chi(x-t)1_{(-\infty,t]} + x^{-1}\chi'(x-t)1_{(-\infty,t)} - t^{-1}\delta_t$, respectively, with point measures δ_t. So, both functions are of bounded variation and their Fourier transforms are bounded by $(1+|u|)^{-1}$ up to multiplicative constants. Finally, observe that g_t is constant and thus Lipschitz near zero, so that g_t is admissible.

For the second claim we again only consider $t < 0$ and first observe, χ being bounded, that

$$\|g_t^c\|_{L^2}^2 \lesssim \int_{-\infty}^{t} |x|^{-2} \sim |t|^{-1}$$

as $t \to -\infty$. Likewise, using the explicit form of $(g_t^c)'$, we see

$$\|g_t^c\|_{H^1} \lesssim \|g_t^c\|_{L^2} + \big\|\big(g_t^c\big)'\big\|_{L^2} \lesssim \big(1+|t|\big)^{-1/2}.$$

For $g_t^s = x^{-1}1_{(-\infty,t]}\chi(x-t)$ we see $\|g_t^s\|_{L^1} \leqslant t^{-1}\|\chi\|_{L^1}$, and the total variation of the derivative of g_t^s is bounded by $t^{-2}\|\chi\|_{L^1} + t^{-1}\|\chi'\|_{L^1} + t^{-1}$. We conclude that $|\mathcal{F}g_t^s(u)| \lesssim (1+|u|)^{-1} \times (1+|t|)^{-1}$ holds. The same argument gives a bound independent of t for $|\mathcal{F}[xg_t^s](u)|$, thus completing the proof. □

Theorem 7. *Suppose Assumption 1 is satisfied, g is admissible and $h_n \sim n^{-1/2}(\log n)^{-\rho}$ for some $\rho > 1$. Then setting*

$$\hat{N}_n(g) := \frac{1}{i\Delta}\int_{\mathbb{R}} g(x)\mathcal{F}^{-1}\big[\big(\varphi_n'/\varphi_n\big)\mathcal{F}K_{h_n}\big](x)\,dx, \qquad N(g) := \int g(x)x\,\nu(dx)$$

(*with some abuse of notation $N(t) = N(g_t)$ etc.*), *we have asymptotic normality,*

$$\sqrt{n}\big(\hat{N}_n(g) - N(g)\big) \to^{\mathcal{L}} N\big(0, \sigma_g^2\big)$$

as $n \to \infty$ with finite variance

$$\sigma_g^2 = (i\Delta)^{-2}\int_{\mathbb{R}} \big(\mathcal{F}^{-1}\big[\mathcal{F}g(u)\varphi^{-1}(-u)\big](x)ix + \mathcal{F}^{-1}\big[\mathcal{F}g(u)\big(\varphi^{-1}\big)'(-u)\big](x)\big)^2\,P(dx).$$

Corollary 8. *Under the assumptions of the preceding theorem the finite-dimensional distributions of the processes $(\sqrt{n}(\hat{N}_n(t) - N(t)),\ t \in \mathbb{R}\setminus\{0\})$ converge to $\mathbb{G}^\varphi$ as $n \to \infty$, where $\mathbb{G}^\varphi$ is a centered Gaussian process, indexed by $\mathbb{R}\setminus\{0\}$, with covariance structure given by (2.6) for $t, s \in \mathbb{R}\setminus\{0\}$.*

Proof. This follows directly by the Cramér–Wold device applied to any finite subfamily of $(g_t,\ t \in \mathbb{R}\setminus\{0\})$, using the preceding lemma and theorem. □

The remaining part of this subsection is devoted to the proof of Theorem 7.

4.1.1. Discarding the drift γ

We shall show that we may assume $\gamma = 0$ in the sequel. To see this, observe that shifting $X_k \mapsto \tilde{X}_k = X_k + \gamma$ leads to the shift in the empirical quotient

$$\varphi_n'(u)/\varphi_n(u) \mapsto \tilde{\varphi}_n'(u)/\tilde{\varphi}_n(u) = \big(e^{iu\gamma}\varphi_n\big)'(u)/\big(e^{iu\gamma}\varphi_n(u)\big) = i\gamma + \varphi_n'(u)/\varphi_n(u)$$

and the true quotient also satisfies $\tilde{\varphi}'(u)/\tilde{\varphi}(u) = i\gamma + \varphi'(u)/\varphi(u)$. In $\hat{N}_n(g) - N(g)$ this shift thus induces the error

$$\begin{aligned}&\left|\frac{1}{i\Delta}\int_{\mathbb{R}} g(x)\mathcal{F}^{-1}\big[i\gamma(\mathcal{F}K_h - 1)\big](x)\,dx\right| \\ &\quad = \frac{|\gamma|}{\Delta}\left|\int_{\mathbb{R}} \big(g(x) - g(0)\big)K_h(x)\,dx\right|\end{aligned}$$

$$\lesssim \int_{\mathbb{R}} \|g\|_{\mathrm{Lip}(0)} |x| \big|K_h(x)\big|\,dx + \int_{[-\delta,\delta]^c} \|g\|_\infty \big|K_h(x)\big|dx$$

$$\lesssim \int_{\mathbb{R}} |x|h^{-1}\big(1 \wedge |x/h|^{-\beta}\big)\,dx + \int_{[-\delta/h,\delta/h]} \big(1+|u|\big)^{-\beta}\,du \lesssim h,$$

where we have used the Lipschitz constant of g in a δ-neighbourhood of zero and (2.4) with $\beta > 2$. By the choice of $h = h_n$ this error is of order $O(h_n) = o(n^{-1/2})$ and thus negligible in the asymptotic distribution of $\sqrt{n}(\hat{N}(g) - N(g))$, and we note that this bound is uniform in all g satisfying the admissibility conditions with uniform constants. Henceforth, without loss of generality, we shall only consider the case $\gamma = 0$.

4.1.2. Approximation error

By approximation error we understand here the deterministic 'bias' term

$$\frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g(-u)\frac{\varphi'(u)}{\varphi(u)}\mathcal{F}K_h(u)\,du - \frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g(-u)\frac{\varphi'(u)}{\varphi(u)}\,du$$

induced by the spectral cutoff with $\mathcal{F}K_h$. We use Assumption 1(b), i.e. that $|\psi'(u)| = |\mathcal{F}[x\nu](u)| \lesssim (1+|u|)^{-1}$. Moreover, we split $g = g^c + g^s$ and treat the bias of each term separately.

For the term involving g^s, using the Lipschitz continuity and boundedness of $\mathcal{F}K$ (due to (2.4) with $\beta > 2$),

$$\begin{aligned}\frac{1}{2\pi\Delta}\left|\int_{\mathbb{R}} \mathcal{F}g^s(-u)\frac{\varphi'(u)}{\varphi(u)}(1-\mathcal{F}K_h)(u)\,du\right| &\lesssim \int_{\mathbb{R}} \big(1+|u|\big)^{-1}\big|\psi'(u)\big|\big|1-\mathcal{F}K(hu)\big|\,du \\ &\lesssim \int_{\mathbb{R}} \big(1+|u|\big)^{-2}\min\big(h|u|,1\big)\,du \\ &\lesssim h\log\big(h^{-1}\big).\end{aligned}$$

For g^c we have by the Cauchy–Schwarz inequality

$$\begin{aligned}\frac{1}{2\pi\Delta}\left|\int_{\mathbb{R}} \mathcal{F}g^c(-u)\frac{\varphi'(u)}{\varphi(u)}(1-\mathcal{F}K_h)(u)\,du\right| &\lesssim \int_{\mathbb{R}} \big(1+|u|\big)\big|\mathcal{F}g^c(-u)\big|\big(1+|u|\big)^{-2}h|u|\,du \\ &\lesssim h\big\|g^c\big\|_{H^1}\left(\int_{\mathbb{R}}\big(1+|u|\big)^{-2}\,du\right)^{1/2} \sim h.\end{aligned}$$

Combining these two estimates, and since $h = h_n = o(n^{-1/2}\log(n)^{-1})$, we conclude that the bias term is of negligible order $o(n^{-1/2})$ in the asymptotic distribution of $\sqrt{n}(\hat{N}(g) - N(g))$.

4.1.3. Main stochastic term

Linearising the error in the quotient φ_n'/φ_n we identify two major stochastic terms:

$$\frac{\varphi_n'(u)}{\varphi_n(u)} - \frac{\varphi'(u)}{\varphi(u)} = \varphi^{-1}(u)\big(\varphi_n' - \varphi'\big)(u) + \big(\varphi^{-1}\big)'(u)(\varphi_n - \varphi)(u) + R_n(u)$$

with remainder

$$R_n(u) := \left(1 - \frac{\varphi_n(u)}{\varphi(u)}\right)\left(\frac{\varphi_n'(u)}{\varphi_n(u)} - \frac{\varphi'(u)}{\varphi(u)}\right) \tag{4.2}$$

where we used the identity $\varphi^{-1}\varphi' + (\varphi^{-1})'\varphi = (\varphi^{-1}\varphi)' = 0$. Discarding the remainder term for the time being, we study the linear centred term

$$\begin{aligned}
&\frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g(-u)\mathcal{F}K_h(u)\big(\varphi^{-1}(u)\big(\varphi_n' - \varphi'\big)(u) + \big(\varphi^{-1}\big)'(u)(\varphi_n - \varphi)(u)\big)\,du \\
&= \frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g(-u)\mathcal{F}K_h(u)\big(\varphi^{-1}(u)\varphi_n'(u) + \big(\varphi^{-1}\big)'(u)\varphi_n(u)\big)\,du \\
&= \frac{1}{2\pi i\Delta}\int_{\mathbb{R}} \mathcal{F}g(-u)\mathcal{F}K_h(u)\big(\varphi^{-1}(u)\mathcal{F}[ixP_n](u) + \big(\varphi^{-1}\big)'(u)\mathcal{F}[P_n](u)\big)\,du \\
&= \frac{1}{i\Delta}\int_{\mathbb{R}} \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g(u)\mathcal{F}K_h(-u)\big](x)ix \\
&\quad + \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g(u)\mathcal{F}K_h(-u)\big](x)\big)\,P_n(dx).
\end{aligned} \tag{4.3}$$

These manipulations are justified by standard Fourier analysis of finite measures, using the compact support of $\mathcal{F}K_h$ and of P_n as well as that $(1+|u|)^{-1}\varphi^{-1}(u)$, $\mathcal{F}g$, $(\varphi^{-1})'$ are all in $L^2(\mathbb{R})$ (by virtue of Assumption 1(c), admissibility of g, Lemma 4(b)).

Thus, the central limit theorem for triangular arrays under Lyapounov's condition (e.g. Theorem 28.3 combined with (28.8) in [1]) applies to the standardised sums if

$$\begin{aligned}
\sup_{h\in(0,1)}\int_{\mathbb{R}} \big|&\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g(u)\mathcal{F}K_h(-u)\big](x)ix \\
&+ \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g(u)\mathcal{F}K_h(-u)\big](x)\big|^{2+\varepsilon}\,P(dx)
\end{aligned} \tag{4.4}$$

is finite.

We use the decomposition $g = g^c + g^s$ and deal with g^c first. We have from the Cauchy–Schwarz inequality, Assumption 1(c) and admissibility of g

$$\int_{\mathbb{R}} \big|\mathcal{F}\big[g^c\big](u)\big|\big|\varphi^{-1}(-u)\big|\,du \leqslant \big\|g^c\big\|_{H^1}\big\|\mathcal{F}^{-1}\big[\varphi^{-1}\big]\big\|_{H^{-1}} < \infty. \tag{4.5}$$

Since also $\sup_{h>0,u} |\mathcal{F}K_h(u)| \leqslant \|K\|_{L^1} < \infty$ we have $\mathcal{F}[g^c]\varphi^{-1}(-\bullet)\mathcal{F}K_h \in L^1(\mathbb{R})$ and thus

$$\sup_{h\in(0,1)} \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g^c(u)\mathcal{F}K_h(-u)\big] \in L^\infty(\mathbb{R}).$$

The integral over the first term in (4.4) with g^c replacing g is thus finite in view of $\int |x|^{2+\varepsilon} P(dx) < \infty$ by Assumption 1(a).

For the singular part we remark $|(\mathcal{F}K_h)'(u)| \leqslant \|xK_h\|_{L^1} \lesssim h$ as well as (by Assumption 1(b)) $|(\varphi^{-1})'(u)| = \Delta|\psi'(u)\varphi^{-1}(u)| \lesssim (1+|u|)^{-1}|\varphi^{-1}(u)|$. We conclude uniformly in h, using admissibility of g,

$$\big|\big(\varphi^{-1}(-\bullet)\mathcal{F}g^s\mathcal{F}K_h(-\bullet)\big)'(u)\big| \lesssim \big|\varphi^{-1}(u)\big|\big(1+|u|\big)^{-1}.$$

By Assumption 1(c) and the Sobolev embedding this implies

$$\sup_h \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g^s(u)\mathcal{F}K_h(-u)\big](x)(1+ix) \in H^\varepsilon(\mathbb{R}) \subseteq L^{2+\varepsilon}(\mathbb{R}). \tag{4.6}$$

Using Lemma 4(a) and $|x|^{2+\varepsilon} \leqslant |x||1+ix|^{2+\varepsilon}$, also the integral over the first term in (4.4) with g^s replacing g is finite.

For the integral over the second term in (4.4) we recall $\sup_{h>0,u} |\mathcal{F}K_h(u)| \leqslant \|K\|_{L^1} < \infty$ and that $\mathcal{F}g, (\varphi^{-1})'$ are both in $L^2(\mathbb{R})$ to deduce $|\mathcal{F}g(u)\mathcal{F}K_h(-u)(\varphi^{-1})'(-u)| \in L^1(\mathbb{R})$ by the Cauchy–Schwarz inequality. By Fourier inversion $\mathcal{F}^{-1}[\mathcal{F}g(u)\mathcal{F}K_h(-u)(\varphi^{-1})'(-u)] \in L^\infty$ holds, and since P is a probability measure, also the integral over the second term is finite.

Altogether we have shown that under our conditions the main stochastic error term is asymptotically normal with rate $1/\sqrt{n}$ and mean zero. For $n \to \infty$ the variances converge to σ_g^2, which follows from $\mathcal{F}K_{h_n} \to 1$ pointwise and uniform integrability by bounded $(2+\varepsilon)$-moments.

4.1.4. Remainder term

In what follows Pr stands for the usual product probability measure $P^{\mathbb{N}}$ describing the joint law of $X_1, X_2, \ldots$, and $Z_n = O_P(r_n)$ means that $r_n^{-1}Z_n$ is bounded in Pr-probability. We show that the remainder term is $O_P(r_n)$ for some $r_n = o(n^{-1/2})$, and therefore negligible in the asymptotic distribution of $\sqrt{n}(\hat{N}(g) - N(g))$.

From Theorem 4.1 of [19] we have for any $\delta > 0$, using the finite $(2+\varepsilon)$-moment property of P from (2.5),

$$\sup_{|u|\leqslant U} \big(\big|\varphi_n(u) - \varphi(u)\big| + \big|\varphi_n'(u) - \varphi'(u)\big|\big) = O_P\big(n^{-1/2}(\log U)^{1/2+\delta}\big).$$

This implies in particular, using

$$\inf_{|u|\leqslant h_n^{-1}} \big|\varphi(u)\big| \gtrsim \inf_{|u|\leqslant h_n^{-1}} \big(1+|u|\big)^{-1/2} \gtrsim \sqrt{h_n} \gtrsim n^{-1/4}(\log n)^{-\rho/2} \tag{4.7}$$

from Lemma 4(b), that for any constant $0 < \kappa < 1$,

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{\varphi_n(u)}\right| < \left|\frac{\kappa}{\varphi(u)}\right| \text{ for some } u \in \left[-h_n^{-1}, h_n^{-1}\right]\right)\\
&\quad = \Pr\left(\left|\frac{\varphi_n(u)}{\varphi(u)}\right| > \kappa^{-1} \text{ for some } u \in \left[-h_n^{-1}, h_n^{-1}\right]\right)\\
&\quad \leqslant \Pr\left(\left|\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}\right| > \left(\kappa^{-1}-1\right) \text{ for some } u \in \left[-h_n^{-1}, h_n^{-1}\right]\right)\\
&\quad \leqslant \Pr\Big(\sup_{|u|\leqslant h_n^{-1}} \left|\varphi_n(u)-\varphi(u)\right| \gtrsim n^{-1/4}(\log n)^{-\rho/2}\Big) \to 0
\end{aligned}$$

as $n \to \infty$, in other words, on events of probability approaching one, φ_n^{-1} decays no faster than φ^{-1} uniformly on increasing sets $[-h_n^{-1}, h_n^{-1}]$.

Now to control the remainder term (4.2) we use $\operatorname{supp}(\mathcal{F}K_h) \subseteq [-h^{-1}, h^{-1}]$ and distinguish each term of the decomposition $g = g^s + g^c$. First, using $|\mathcal{F}g^s(u)| \lesssim (1+|u|)^{-1}$, Lemma 4(b) and Assumption 1(c) we see

$$\begin{aligned}
&\left|\int_{-h^{-1}}^{h^{-1}} \mathcal{F}g^s(-u)\mathcal{F}K_h(u)R_n(u)\,du\right|\\
&\quad = O_P\left(\int_{-h^{-1}}^{h^{-1}} \left(1+|u|\right)^{-1} n^{-1}\left(\log h^{-1}\right)^{1+2\delta}\left|\varphi^{-1}(u)\right|\left(\left|\varphi(u)^{-1}\right| + \left|\left(\varphi^{-1}\right)'(u)\right|\right)du\right)\\
&\quad = O_P\left(n^{-1}\left(\log h^{-1}\right)^{1+2\delta} h^{2\varepsilon-1}\int \left(1+|u|\right)^{-2+2\varepsilon}\left|\varphi(u)\right|^{-2}du\right)\\
&\quad = O_P\left(n^{-1}\left(\log h^{-1}\right)^{1+2\delta} h^{2\varepsilon-1}\right).
\end{aligned}$$

For the nonsingular part we have likewise, using the Cauchy–Schwarz inequality, $g^c \in H^1(\mathbb{R})$, (4.7) and Assumption 1(c),

$$\begin{aligned}
&\left|\int_{-h^{-1}}^{h^{-1}} \mathcal{F}g^c(-u)\mathcal{F}K_h(u)R_n(u)\,du\right|\\
&\quad = O_P\left(n^{-1}\left(\log h^{-1}\right)^{1+2\delta}\left(\int_{-h^{-1}}^{h^{-1}} \left(1+|u|\right)^{-2}\left|\varphi(u)\right|^{-4}du\right)^{1/2}\right)\\
&\quad = O_P\left(n^{-1}\left(\log h^{-1}\right)^{1+2\delta} h^{-1/2}\left\|\varphi^{-1}\left(1+|u|\right)^{-1}\right\|_{L^2}\right).
\end{aligned}$$

Consequently, the remainder term is negligible because $h_n^{-1+2\varepsilon}(\log h_n^{-1})^{1+2\delta} = o(n^{1/2})$. Note that this gives in fact uniform $o_P(n^{-1/2})$-control of the remainder term for all g that satisfy the admissibility bounds uniformly.

4.2. Tightness of the linear term

We study the linear part (4.3) and introduce the empirical process

$$\nu_n^{\varphi}(t) := \sqrt{n}\frac{1}{i\Delta}\int_{\mathbb{R}} \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t(u)\mathcal{F}K_{h_n}(-u)\big](x)ix \\ + \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g_t(u)\mathcal{F}K_{h_n}(-u)\big](x)\big)(P_n - P)(dx), \quad |t| \geqslant \zeta > 0. \tag{4.8}$$

Recall that this process is centered even without subtracting P. Moreover, since $\sup_{|t|\geqslant\zeta}\|g_t\|_{L^2} < \infty$, the arguments after (4.3) imply that ν_n^{φ} is a (possibly non-measurable) random element of the space $\ell^{\infty}((-\zeta,\zeta)^c)$ of bounded functions on $(-\infty,-\zeta]\cup[\zeta,\infty)$ (the complement of $(-\zeta,\zeta)$ in $\mathbb{R}$) equipped with the uniform norm $\|\bullet\|_{(-\zeta,\zeta)^c}$.

4.2.1. Pregaussian limit process

Theorem 2 will follow if we show that ν_n^{φ} converges to $\mathbb{G}^{\varphi}$ in law in $\ell^{\infty}((-\zeta,\zeta)^c)$. For this statement to make sense we have to show first that $\mathbb{G}^{\varphi}$ defines a proper Borel random variable in $\ell^{\infty}((-\zeta,\zeta)^c)$, which is implied by the following more general result. Recall that any Gaussian process $\{\mathbb{G}(t)\}_{t\in T}$ induces its intrinsic covariance metric $d^2(s,t) = E(\mathbb{G}(s) - \mathbb{G}(t))^2$ on the index set T.

Theorem 9. *Grant Assumption 1. The Gaussian process $\{\mathbb{G}^{\varphi}(t)\}_{t:\,|t|\geqslant\zeta}$ with covariance given by (2.6) admits a version, still denoted by $\mathbb{G}^{\varphi}$, which has uniformly continuous sample paths almost surely for the intrinsic covariance metric of $\mathbb{G}^{\varphi}$, and which satisfies* $\sup_{t:\,|t|\geqslant\zeta}|\mathbb{G}^{\varphi}(t)| < \infty$ *almost surely.*

The proof moreover implies that $(-\zeta,\zeta)^c$ is totally bounded in the metric d. Therefore (a version of) $\mathbb{G}^{\varphi}$ concentrates on the separable subspace of $\ell^{\infty}((-\zeta,\zeta)^c)$ consisting of bounded d-uniformly continuous functions on $(-\zeta,\zeta)^c$, from which we may in particular conclude that $\mathbb{G}^{\varphi}$ defines a Borel-random variable in that space, and hence is also a Borel random variable in the ambient space $\ell^{\infty}((-\zeta,\zeta)^c)$.

Next to Dudley's entropy integral, the main tool in the proof of Theorem 9 is the following bound for the pseudo-differential operator $\mathcal{F}^{-1}[\varphi^{-1}(-u)]$. For $f \in L^2(\mathbb{R})$ we set $\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)] * f := \mathcal{F}^{-1}[\varphi^{-1}(-u)\mathcal{F}f(u)]$ which is well defined at least in $H^{(1-\varepsilon)/2}(\mathbb{R})$ in view of Lemma 4. Alternatively, $\|\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)]*f\|_{L^2} \lesssim \|(1+|u|)^{(1-\varepsilon)/2}\mathcal{F}f(u)\|_{L^2}$ whenever $f \in H^{(1-\varepsilon)/2}(\mathbb{R})$, but such an inequality is not sufficient for our purposes. We need a stronger estimate for functions f supported away from the origin, and with the $\|\bullet\|_{L^2}$-norm replaced by the $\|\bullet\|_{2,P}$-norm. Intuitively speaking, and considering the example $f = 1_{(s,t]}$, $s < t < 0$, relevant below, this strengthening is possible since the locations of singularities of $1_{(s,t]}$ and of P (at the origin) are separated away from each other, and since this remains so after application of the pseudo-local operator $\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)] * (\bullet)$ to f.

Proposition 10. *Grant Assumption 1 and define $\|h\|_{2,P} := (\int h^2\,dP)^{1/2}$. For $f \in L^2(\mathbb{R})$ with* $\operatorname{supp}(f)\cap(-\delta,\delta) = \varnothing$ *for some $\delta > 0$ we have*

$$\begin{aligned} &\big\|\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big]*f\big\|_{2,P} \\ &\quad\lesssim \big\|\big(1+|u|\big)^{1-\varepsilon}\mathcal{F}f(u)\big\|_{L^{2+4/\varepsilon}(\mathbb{R})}+\bigg(\int\frac{f(y)^2}{1+y^2}\,dy\bigg)^{1/2} \end{aligned} \tag{4.9}$$

provided the right-hand side is finite. The constant in this bound depends only on δ.

Proof. We shall need the pseudo-differential operator identity

$$\begin{aligned} &\big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big]*f\big)(x) \\ &\quad=\bigg(\bigg(\frac{1}{i\bullet}\mathcal{F}^{-1}\big[\big(\varphi^{-1}(-u)\big)'\big]\bigg)*f\bigg)(x),\quad f\in L^2(\mathbb{R}),\ x\notin\operatorname{supp}(f), \end{aligned} \tag{4.10}$$

where the right-hand side is defined classically. This identity is fundamental for establishing the property of pseudo-locality in a C^∞-framework, see e.g. Theorems 8.8 and 8.9 in [9]. Let us verify this identity here, where $\varphi^{-1}\notin C^\infty$. Consider $f\in L^2(\mathbb{R})$ and g any smooth compactly supported test function such that $\operatorname{supp}(f)\cap\operatorname{supp}(g)=\varnothing$. Then $(f*g(-\bullet))(0)=0$ and $f*g$ is smooth from which we may conclude that also $x^{-1}(f*g(-\bullet))(x)$ (equal to $(f*g)'(0)$ at zero) is in $L^2(\mathbb{R})$ and smooth, and that

$$\mathcal{F}\bigg[\frac{(f*g(-\bullet))(x)}{ix}\bigg]'(u)=\mathcal{F}\big[f*g(-\bullet)\big](u)=\mathcal{F}f(u)\overline{\mathcal{F}g(u)}.$$

Plancherel's formula, integration by parts and Fubini's theorem (using $(\varphi^{-1})'\in L^2(\mathbb{R})$ from Lemma 4 and the support properties) yield

$$\begin{aligned} \int\big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big]*f\big)(x)g(x)\,dx&=\frac{1}{2\pi}\int\varphi^{-1}(-u)\mathcal{F}f(u)\overline{\mathcal{F}g(u)}\,du \\ &=\frac{-1}{2\pi}\int\big(\varphi^{-1}(-u)\big)'\mathcal{F}\bigg[\frac{(f*g(-\bullet))(x)}{ix}\bigg](u)\,du \\ &=\int\frac{\mathcal{F}^{-1}[(\varphi^{-1}(-u))'](x)}{ix}\big(f*g(-\bullet)\big)(-x)\,dx \\ &=\int\bigg(\frac{\mathcal{F}^{-1}[(\varphi^{-1}(-u))']}{i\bullet}*f\bigg)(x)g(x)\,dx. \end{aligned}$$

In this calculation the boundary terms vanish due to the fast decay of $\mathcal{F}[(f*g(-\bullet))(x)/x]$ (g smooth). Consequently, (4.10) follows by testing with all g supported near x.

We use Hölder's inequality, the Hausdorff–Young inequality from Fourier analysis, the bound $p(x)\lesssim|x|^{-1}$ from Lemma 4, the pseudo-differential operator identity, again Hölder's inequality, Assumption 1(c) and $(\varphi^{-1})'\in L^2$ in view of Lemma 4 in this order to obtain for $\delta'=\delta/2$:

$$\begin{aligned} &\bigg|\int\big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big]*f\big)^2(x)\,P(dx)\,dx\bigg| \\ &\quad\leqslant\big\|\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big]*f\big\|^2_{L^{2+\varepsilon}(\mathbb{R})}\|p\|_{L^{(2+\varepsilon)/\varepsilon}([-\delta',\delta']^c)} \end{aligned}$$

$$
\begin{aligned}
&\quad + \big\|\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * f\big\|^2_{L^\infty([-\delta',\delta'])} P\big([-\delta',\delta]\big) \\
&\lesssim \big\|\varphi^{-1}(-u)\mathcal{F}f\big\|^2_{L^{(2+\varepsilon)/(1+\varepsilon)}} \|xp\|_\infty \big(\delta'\big)^{-2/(2+\varepsilon)} \\
&\quad + \big\|\big(\mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\big](x)/x\big) * f\big\|^2_{L^\infty([-\delta',\delta'])} \\
&\lesssim \big\|\varphi^{-1}(-u)\big(1+|u|\big)^{-1+\varepsilon}\big\|^2_{L^2} \big\|\big(1+|u|\big)^{1-\varepsilon}\mathcal{F}f(u)\big\|^2_{L^{2+4/\varepsilon}} \big(\delta'\big)^{-2/(2+\varepsilon)} \\
&\quad + \big\|\mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'\big]\big\|^2_{L^2} \sup_{x\in[-\delta',\delta']} \int \frac{f(y)^2}{(x-y)^2}\,dy \\
&\lesssim \big\|\big(1+|u|\big)^{1-\varepsilon}\mathcal{F}f(u)\big\|^2_{L^{2+4/\varepsilon}} + \int \frac{f(y)^2}{1+y^2}\,dy,
\end{aligned}
$$

provided f is such that the last line is finite. Take square roots to deduce the asserted inequality with a constant independent of f. □

Proof of Theorem 9. We consider the generalised Brownian bridge process arising as the pointwise weak limit of (4.8), so with $\mathcal{F}K_h \equiv 1$, and further split $g_t = g_t^c + g_t^s$ as in the proof of Lemma 6. More precisely, we study the Gaussian process indexed by $(i\Delta)^{-1}$ times

$$
\begin{aligned}
h_t(x) &= \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g_t(u)\big](x) + \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t(u)\big](x)\big)ix \\
&= \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g_t(u)\big](x) \\
&\quad + \big(\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t^c(u)\big](x) + \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t^s(u)\big](x)\big)ix,
\end{aligned}
\tag{4.11}
$$

where $|t| \geqslant \zeta$. The theorem is thus proved if we show that the class of functions $\mathcal{G} = \{(i\Delta)^{-1}h_t\colon t \in \mathbb{R}\setminus(-\zeta,\zeta)\}$ is bounded in $L^2(P)$ and P-pregaussian (cf. [8, Chapter 2, pp. 92–93]). In Section 4.1.3 above we have shown the $L^{2+\varepsilon}(P)$-boundedness of the same function class, but also involving the kernel K_h. The same proof, replacing $\mathcal{F}K_h$ just by one, shows that $\mathcal{G}$ is even $L^{2+\varepsilon}(P)$-bounded. To establish that $\mathcal{G}$ is pregaussian it suffices, by Dudley's integral-criterion, to find a suitable η-covering of $\mathcal{G}$ in the intrinsic covariance metric $d(s,t) := \|h_t - h_s\|_{2,P}$, for every $h_t, h_s \in \mathcal{G}$.

Consider first increments for $s < t$, $|s-t| \leqslant 1$, $\min(|s|,|t|) \geqslant \zeta$,

$$
\begin{aligned}
h_t(x) - h_s(x) &= \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}[g_t - g_s](u)\big](x) + \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}[g_t - g_s](u)\big](x)ix \\
&= i\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}\big[x(g_t - g_s)\big](u)\big](x) \\
&= i\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\big] * 1_{(s,t]}(x),
\end{aligned}
$$

for which Proposition 10 yields, with $f = 1_{(s,t]}$, the Hölder-type bound

$$
\|h_t - h_s\|^2_{2,P} \lesssim \big\|\sin\big((t-s)u\big)u^{-1}\big(1+|u|\big)^{1-\varepsilon}\big\|^2_{L^{2+4/\varepsilon}} + |t-s| \lesssim |t-s|^{\varepsilon(3+2\varepsilon)/(2+\varepsilon)}.
$$

This will give us a polynomially growing covering of $\mathcal{G}$ for all t in a fixed compact interval.

To deal with large $|t|$ we shall establish the polynomial decay bound $\|h_t\|_{2,P} \lesssim |t|^{-1/2}$ as $|t| \to \infty$, and we shall do this for each of the three terms in the second line of (4.11) separately.

For the first term, say $h_t^{(1)}$, this follows from

$$\|h_t^{(1)}\|_{2,P}^2 \leqslant \|h_t^{(1)}\|_{L^\infty}^2 \leqslant \|(\varphi^{-1})'(-\bullet)\mathcal{F}g_t\|_{L^1}^2 \leqslant \|(\varphi^{-1})'(-\bullet)\|_2^2\|g_t\|_{L^2}^2 \lesssim \int_{-\infty}^{t} |x|^{-2} \sim |t|^{-1}$$

as $t \to -\infty$, and likewise for $t \to \infty$, using the Cauchy–Schwarz inequality and Lemma 4(b).

For the second term $h_t^{(2)}$ we use the Cauchy–Schwarz inequality, the finite second moment of P, Assumption 1(c) and Lemma 6 to the effect that

$$\|h_t^{(2)}\|_{2,P}^2 \leqslant \int x^2\, P(dx)\|\varphi^{-1}(-u)\mathcal{F}g_t^c(u)\|_{L^1}^2 \lesssim \|\mathcal{F}^{-1}[\varphi^{-1}]\|_{H^{-1}}^2\|g_t^c\|_{H^1}^2 \lesssim (1+|t|)^{-1}.$$

For the third term, since xP has a bounded density by Lemma 4(a), it suffices to bound

$$\|\mathcal{F}^{-1}[\varphi^{-1}(-u)\mathcal{F}g_t^s](x)|x|^{1/2}\|_{L^2},$$

which by the Cauchy–Schwarz inequality can be estimated by

$$\|\mathcal{F}^{-1}[\varphi^{-1}(-u)\mathcal{F}g_t^s]x\|_{L^2}^{1/2}\|\mathcal{F}^{-1}[\varphi^{-1}(-u)\mathcal{F}g_t^s]\|_{L^2}^{1/2}.$$

Now by Lemma 6 we know $|\mathcal{F}g_t^s(u)| \lesssim (1+|u|)^{-1}(1+|t|)^{-1}$, $|\mathcal{F}[xg_t^s](u)| \lesssim (1+|u|)^{-1}$ and since $|(\varphi^{-1})'|(u) \lesssim (1+|u|)^{-1}|\varphi^{-1}(u)|$ from the proof of Lemma 4 we can estimate the product in the last display to obtain the overall bound

$$\|h_t^{(3)}\|_{2,P} \lesssim (1+|t|)^{-1/2}\|\varphi^{-1}(-u)(1+|u|)^{-1}\|_{L^2} \lesssim (1+|t|)^{-1/2}$$

in view of Assumption 1(c).

In conclusion, we can construct an η-covering of $\mathcal{G}$ by the functions $(i\Delta)^{-1}h_{t_i}$ with $t_i = i/M$ and $i = -M^2, \dots, +M^2$ where $M = M(\eta)$ grows polynomially in η^{-1}. This shows that the covering numbers corresponding to this η-net satisfy

$$\log(N(\mathcal{G}, L^2(P), \eta)) \lesssim \log(\eta^{-1}). \tag{4.12}$$

The square-root of this entropy bound is integrable at zero as a function of η, which completes the proof by Dudley's continuity criterion (Theorem 2.6.1 in [8]). □

4.2.2. Uniform CLT for the linear term

Theorem 11. *Grant Assumption 1 and*

$$(\nu_n^\varphi(t_1), \dots \nu_n^\varphi(t_k)) \to^{\mathcal{L}} (\mathbb{G}^\varphi(t_1), \dots, \mathbb{G}^\varphi(t_k))$$

as $n \to \infty$ for every finite set $(t_1, \dots, t_k) \subseteq (-\zeta, \zeta)^c$. If $h_n \gtrsim n^{-1/(4\alpha)}$ for some $\alpha > (1-\varepsilon)/2$, so in particular if $h_n \sim n^{-1/2}(\log n)^{-\rho}$ for some $\rho > 1$, then

$$\nu_n^\varphi \to^{\mathcal{L}} \mathbb{G}^\varphi \quad \text{in } \ell^\infty((-\zeta, \zeta)^c)$$

as $n \to \infty$.

Proof. We set $\Delta = 1$ and suppose that the kernel is symmetric, i.e. $\mathcal{F}K_h(-u) = \mathcal{F}K_h(u)$, to ease notation. Given convergence of the finite-dimensional distributions it suffices to prove uniform tightness of $\{\nu_n^\varphi\}_{n\in\mathbb{N}}$ in $\ell^\infty((-\zeta,\zeta)^c)$, cf. [27, Chapter 1.5]. We shall in what follows decompose ν_n^φ into a sum of several processes indexed by t, and prove tightness of each of these processes separately, which implies tightness of the sum of the processes by the asymptotic equicontinuity characterisation of tightness in $\ell^\infty((-\zeta,\zeta)^c)$ (e.g., Theorem 1.5.7 in [27]) and by the triangle inequality. We shall also frequently use the simple fact that tightness is preserved under isometric injections of $\ell^\infty((-\zeta,\zeta)^c)$: if ν is a process indexed by s and ν' a process indexed by functions $f_s \in \mathcal{F}$, and if $\nu(s) = \nu'(f_s)$ for every $s \in (-\zeta,\zeta)^c$, then tightness of ν' in $\ell^\infty(\mathcal{F})$ (normed by $\|H\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |H(f)|$) implies tightness of ν in $\ell^\infty((-\zeta,\zeta)^c)$.

We decompose $g_t = g_t^c + g_t^s$ as in the proof of Lemma 6 with the particular choice $\chi(x) = e^x 1_{(-\infty,0]}(x)$ for $t < 0$, and symmetrically if $t > 0$. The integrand of $\nu_n^\varphi(t)$ in (4.8) equals

$$\begin{aligned}
&\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}\big[ixg_t^s\big]\mathcal{F}K_h\big](x) + \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t^s\mathcal{F}[ixK_h]\big](x)\\
&\quad + \mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t^c\mathcal{F}K_h\big](x)ix + \mathcal{F}^{-1}\big[\big(\varphi^{-1}\big)'(-u)\mathcal{F}g_t^c\mathcal{F}K_h\big](x)\\
&=: (T_1 + T_2 + T_3 + T_4)(x).
\end{aligned}$$

The process indexed by the component T_1 is critical and its tightness is proved in Section 4.2.3 below.

Concerning T_2, we have $|\varphi^{-1}(-u)\mathcal{F}g_t^s\mathcal{F}[ixK_h]| \lesssim |\varphi^{-1}(-u)|(1+|u|)^{-2}$ by $\|xK_h\|_{L^1} + \|(xK_h)'\|_{L^1} \lesssim 1$, uniformly in h, and by the admissibility of g_t. By Assumption 1(c) we deduce that T_2 lies in a fixed norm ball of $H^1(\mathbb{R})$. For T_4 we note $|(\varphi^{-1})'| \lesssim 1$, $\sup_{h>0,u} |\mathcal{F}K_h(u)| \leqslant \|K\|_1 < \infty$, $\sup_{|t|\geqslant\zeta} \|g_t^c\|_{H^1} < \infty$ by Lemmas 4 and 6, so $\{\mathcal{F}^{-1}[(\varphi^{-1})'(-u)\mathcal{F}g_t^c\mathcal{F}K_h](\bullet), |t| \geqslant \zeta\}$ is bounded in $H^1(\mathbb{R})$. For T_3 we use $|\varphi^{-1}(u)| \leqslant (1+|u|)^{(1-\varepsilon)/2}$ and

$$\big\|\mathcal{F}^{-1}\big[\varphi^{-1}(-u)\mathcal{F}g_t^c\mathcal{F}K_h\big]\big\|_{H^{(1+\varepsilon)/2}} \lesssim \big\|\big(1+|u|\big)\mathcal{F}g_t^c\big\|_{L^2} = \|g_t^c\|_{H^1} < \infty,$$

uniformly in $|t| \geqslant \zeta$, again by Lemmas 4 and 6. We conclude that the norms $\|T_2 + T_4\|_{H^1}$ and $\|T_3/x\|_{H^{(1+\varepsilon)/2}}$ are bounded uniformly in $t \in (-\zeta,\zeta)^c$, $h > 0$. Each summand in $T_2 + T_3 + T_4$ is therefore contained in a fixed P-Donsker-class: For $T_2 + T_4$ this follows from Proposition 1 in [21] with $s = 1$, $p = q = 2$, and for T_3 we apply Corollary 5 for weighted Besov–Sobolev spaces in [21] with parameter choice $s = (1+\varepsilon)/2$, $\beta = -1$, $p = q = 2$, $\gamma = \varepsilon/2$ noting that the moment condition there is satisfied by (2.5). The empirical process ν_n^φ is thus indexed by functions $T_2 + T_3 + T_4$ that change with n but that are contained in a fixed P-Donsker class, and so is tight by the asymptotic equicontinuity criterion. Together with the tightness of the critical term, derived below, this proves tightness of ν_n^φ. □

Combining the convergence of the finite-dimensional distributions from Corollary 8 with Theorem 11 and the uniform bounds on the remainder and bias term we have succeeded in proving Theorem 2.

4.2.3. The critical term

Note that in the ill-posed case $\lim_{|u|\to\infty} |\varphi(u)| = 0$, for instance when $\varphi(u) = (1-iu)^{-\alpha}$, the class involving T_1 with $\mathcal{F}K_h = 1$ *is not* P-Donsker even for P with bounded density. The reason is, roughly speaking, that $\mathcal{F}^{-1}[\varphi^{-1}(-\bullet)] * (e^{\bullet - t}1_{(-\infty,t]})$ is then unbounded at t, and classes that contain functions unbounded at any point cannot be Donsker for such P, cf. the proof of Theorem 7 in [20]. This implies that one cannot use $h = 0$, i.e., $K_h = \delta_0$, in the proofs, as could have been done in the 'noncritical' terms T_2, T_3, T_4 above. Rather, one needs to exploit the fact that the kernel K_h smooths out the singularities for h fixed, and if h_n does not approach zero too fast, there is still hope to obtain a uniform central limit theorem, as shown in a different but conceptually related situation of Theorems 9 and 10 in [10].

As compactly supported kernels facilitate the arguments considerably, we introduce the truncated kernel

$$K_h^{(0)} := K_h \mathbf{1}_{[-\zeta/2,\zeta/2]}.$$

By the decay of K and K' from (2.4) we can again treat the term involving $K_h - K_h^{(0)}$ by classical methods. Using $\|K_h - K_h^{(0)}\|_{BV} \lesssim h^{\beta-2}$ where $\|\bullet\|_{BV}$ is the usual bounded variation norm, we obtain

$$\left|\varphi^{-1}(-u)\mathcal{F}\left[ixg_t^s\right]\mathcal{F}\left[K_h - K_h^{(0)}\right](u)\right| \lesssim \left|\varphi^{-1}(-u)\right|\left(1+|u|\right)^{-2}h^{\beta-2},$$

whence $\mathcal{F}^{-1}[\varphi^{-1}(-u)\mathcal{F}[ixg_t^s]\mathcal{F}[K_h - K_h^{(0)}]] \in H^1(\mathbb{R})$ follows, even with in h shrinking and in t uniform norms. As for the terms T_2, T_4 above, we thus deduce the uniform tightness of this term since norm balls in $H^1(\mathbb{R})$ are universally Donsker.

Recalling $g_t^s(x) = x^{-1}e^{x-t}\mathbf{1}_{(-\infty,t]}(x)$, the term involving the truncated kernel can be written as

$$\mathcal{F}^{-1}\left[\varphi^{-1}(-u)\mathcal{F}\left[ixg_t^s\right]\mathcal{F}K_h^{(0)}\right] = iq(\bullet - t) * K_h^{(0)}$$

with

$$q(x) := \mathcal{F}^{-1}\left[\varphi^{-1}(-u)(1+iu)^{-1}\right](x). \tag{4.13}$$

The regularity of q in the scale of Besov spaces $B^s_{p,r}(\mathbb{R})$ is $s = (1+\varepsilon)/2$ for $p = 1$ and $r = \infty$: Since $m(u) = \varphi^{-1}(-u)(1+iu)^{-1/2+\varepsilon/2}$ is a Fourier multiplier on $B^{(1+\varepsilon)/2}_{1,\infty}(\mathbb{R})$ by Lemma 4(c), this assertion follows from the fact that

$$\mathcal{F}^{-1}\left[(1+iu)^{-1/2-\varepsilon/2}\right](x) = \Gamma(1/2+\varepsilon/2)^{-1}|x|^{\varepsilon/2-1/2}e^x\mathbf{1}_{(-\infty,0]}(x)$$

(a Gamma-type density) is an element of that space. The latter follows either by checking directly that its L^1-modulus of smoothness satisfies $\omega(h)_1 \lesssim h^{1/2+\varepsilon}$ or by noting that multiplication by $(1+iu)^{(1-\varepsilon)/2}$ in the Fourier domain is an isomorphism between $B^{(1+\varepsilon)/2}_{1,\infty}(\mathbb{R})$ and $B^1_{1,\infty}(\mathbb{R})$ and $\mathcal{F}^{-1}[(1+iu)^{-1}](x) = e^x\mathbf{1}_{(-\infty,0]}(x)$ is of bounded variation and thus contained in $B^1_{1,\infty}(\mathbb{R})$. Moreover, by embedding theorems for Besov spaces, q is then also an element of $B^s_{1,1}(\mathbb{R})$ for any $s < (1+\varepsilon)/2$ and thus also of $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. We refer to [26] for these standard properties of Besov spaces.

We are thus left with proving tightness of

$$\sqrt{n}\int_{\mathbb{R}}\big(q(\bullet-t)*K_h^{(0)}\big)(x)\,(P_n-P)(dx)$$
$$=\sqrt{n}\int_{\mathbb{R}} q(y-t)\big(K_h^{(0)}*(P_n-P)\big)(y)\,dy,\quad |t|\geqslant\zeta, \tag{4.14}$$

which is a smoothed empirical process indexed by

$$\mathcal{F}=\big\{q(\bullet-t)\colon\ |t|\geqslant\zeta\big\}. \tag{4.15}$$

The following general purpose result follows from the proof of Theorem 3 in [10], which builds on fundamental ideas in the classical paper [12], and can be applied to the unbounded processes relevant here. For a given class of measurable functions $\mathcal{F}$ we write

$$\mathcal{F}'_\delta=\big\{f-g\colon\ f,g\in\mathcal{F},\ \|f-g\|_{2,P}\leqslant\delta\big\}.$$

We shall rather loosely use the standard empirical process terminology from [10].

Theorem 12. *Let $\mathcal{F}$ be any P-pregaussian class of real-valued functions on $\mathbb{R}^d$ and let $\{\mu_n\}_{n=1}^\infty$ be a sequence of finite signed measures defined on $\mathbb{R}^d$ satisfying $\sup_n\|\mu_n\|<\infty$. Let $\bar\mu_n(A)=\mu_n(-A)$. Assume that $\mathcal{F}\subseteq L^1(|\mu_n|)$ holds for every n and, in addition,*

(a) *for each n, the class $\tilde{\mathcal{F}}_n:=\{f*\bar\mu_n\colon\ f\in\mathcal{F}\}$ consists of functions whose absolute values are bounded by a constant M_n;*
(b) $\sup_{f\in\mathcal{F}'_\delta}E(f*\bar\mu_n(X))^2\leqslant 4\delta^2$ *for every $\delta>0$ and $n\geqslant n_0\equiv n_0(\delta)$ large enough;*
(c) *for i.i.d. Rademacher variables $(\varepsilon_i)_i$, independent of the X_i's, we have*

$$\Bigg\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i f(X_i)\Bigg\|_{(\tilde{\mathcal{F}}_n)'_{1/n^{1/4}}}\to 0 \tag{4.16}$$

as $n\to\infty$ in outer probability;
(d) $\bigcup_{n\geqslant 1}\tilde{\mathcal{F}}_n$ *is in the $L^2(P)$-closure of* $\sup_n\|\mu_n\|$*-times the symmetric convex hull of some fixed P-pregaussian class of functions $\bar{\mathcal{F}}$;*
(e) *for all $0<\eta<1$, the $L^2(P)$-metric entropy of $\tilde{\mathcal{F}}_n$ satisfies $H(\tilde{\mathcal{F}}_n,L^2(P),\eta)\leqslant\lambda_n(\eta)/\eta^2$ for functions $\lambda_n(\eta)$ such that $\lambda_n(\eta)\to 0$ and $\lambda_n(\eta)/\eta^2\to\infty$ as $\eta\to 0$, uniformly in n, and the bounds M_n of part* (a) *satisfy*

$$M_n\leqslant\Big(5\sqrt{\lambda_n\big(1/n^{1/4}\big)}\Big)^{-1} \tag{4.17}$$

for all n large enough.

Then $\sqrt{n}(P_n-P)*\mu_n$ is uniformly tight in the Banach space $\ell^\infty(\mathcal{F})$ (equipped with the uniform norm $\|\bullet\|_{\mathcal{F}}$).

Proof. The differences to Theorem 3 in [10] are: We do not require $\mu_n(\mathbb{R}) = 1$, $\forall n$, and (b) is slightly weakened, both permitted as we only establish tightness in this theorem and not convergence of the finite-dimensional distributions. Moreover the new condition (d), which replaces translation invariance of $\mathcal{F}$ by a more generic condition. Note that Theorem 0.3 in [7] implies that $L^2(P)$-closures of symmetric convex hulls of pregaussian classes are again pregaussian, which is all that is needed for the proof of Theorem 3 in [10] to apply. □

We now verify these conditions for the classes above, with $d\mu_n(y) = K_h^{(0)}(y)\,dy$. Let us first show that the class $\mathcal{F}$ from (4.15) is indeed P-pregaussian. By Proposition 10 applied to

$$f(x) = e^x\big(e^{-t}1(x \leqslant t) - e^{-s}1(x \leqslant s)\big), \quad t, s \leqslant -\zeta \text{ (and symmetrically for } t, s \geqslant \zeta)$$

and by the same estimates as in the proof of Theorem 9

$$\big\| q(\bullet - t) - q(\bullet - s)\big\|_{2,P} \lesssim |t - s|^{\varepsilon(3+2\varepsilon)/(2+2\varepsilon)}. \tag{4.18}$$

Moreover, the tail bound for the third term in that proof applies here such that the same arguments show that $\mathcal{F}$ has polynomially growing covering numbers and is thus pregaussian. In particular, $\mathcal{F}$ is bounded in $L^2(P)$. The functions $q(\bullet - t)$ are in $B^{(1+\varepsilon)/2}_{1,\infty}(\mathbb{R}) \subseteq L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and thus in $L^1(|\mu_n|)$ since K is bounded.

(a) The envelopes of $q(\bullet - t) * K_h^{(0)}$ are of order $M_n \lesssim h^{-\alpha'}$ for $\alpha' \in ((1-\varepsilon)/2, \alpha)$ when $h = h_n \gtrsim n^{-1/(4\alpha)}$ since the sup-norm is bounded by the BV-norm, which in turn is bounded in point (c) below.

(b) Let $g \in \mathcal{F}'_\delta$, then $\|K_h^{(0)} * g\|_{2,P} \leqslant \|K_h^{(0)} * g - g\|_{2,P} + \delta$ and the result follows from the triangle inequality if we show $\|K_h^{(0)} * f - f\|_{2,P} \to 0$ uniformly over $f \in \mathcal{F}$. From (4.9) above, noting $\operatorname{supp}(K_h^{(0)} * (i\bullet g_t^s)) \cap (-\zeta/2, \zeta/2) = \varnothing$, we conclude

$$\big\| K_h^{(0)} * f - f\big\|_{2,P} \lesssim \big\| (1+|u|)^{-\varepsilon}\big(\mathcal{F}K_h^{(0)} - 1\big)\big\|_{L^{2+4/\varepsilon}} + \big\| K_h^{(0)} * f - f\big\|_{L^2}.$$

Since $\mathcal{F}K_h^{(0)}(u)$ is uniformly bounded and tends to 1 pointwise and since $(1+|u|)^{-(2\varepsilon+4)}$ is integrable, by dominated convergence the first norm tends to zero for $h \to 0$. Similarly, as $\mathcal{F}f \in L^2$ holds, $\|(\mathcal{F}K_h^{(0)} - 1)\mathcal{F}f\|_{L^2} \to 0$ follows and by Plancherel's theorem also the second norm converges to zero. This convergence is uniform because of $|\mathcal{F}f(u)| = |\mathcal{F}q(u)|$ for all $f \in \mathcal{F}$ and since $q \in L^2(\mathbb{R})$.

(c) The class $\{K_h^{(0)} * q(\bullet - t)\colon |t| \geqslant \zeta\}$ consist of translates of the fixed function $K_h^{(0)} * q$, which is a function of bounded variation with BV-norm of size $h^{-\alpha'}$ for some $\alpha' \in ((1-\varepsilon)/2, \alpha)$ using $q \in B^{1-\alpha'}_{11}(\mathbb{R})$ from the argument after (4.13) and the estimate (61) in [10] (whose proof applies also to the truncated kernels). The envelope M_n of $\tilde{\mathcal{F}}_n$ is then of the same size since the BV-norm bounds the supremum norm. Moreover the class $\{K_h^{(0)} * q(\bullet - t)\colon |t| \geqslant \zeta\}$ has polynomial $L^2(Q)$-covering numbers, uniformly in all probability measures Q. To see this we argue as in Lemma 1 in [11]: note that a function of bounded variation is the composition of a 1-Lipschitz function with a monotone function. The set of all translates of a monotone function has VC-index 2, and hence has polynomial covering numbers by Theorem 5.1.15 in [6], with constants

A, v there independent of n. Composition with a 1-Lipschitz map preserves the entropy, and the estimate (22) in [10] with envelopes $M_n \sim h_n^{-\alpha'}$ and $H_n(\eta) \equiv H(\eta) \sim \log(\eta)$ now shows that

$$E\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{(\tilde{\mathcal{F}}_n)'_{1/n^{1/4}}} \lesssim \max\left[\frac{\sqrt{\log n}}{n^{1/4}}, \frac{h_n^{-\alpha'}}{\sqrt{n}} \log n \right] \to 0$$

as $n \to \infty$, in view of $h_n^{-\alpha'} \lesssim h_n^{-\alpha} \lesssim n^{1/4}$.

(d) Using that $K_h^{(0)}$ is supported in $[-\zeta/2, \zeta/2]$, one shows by standard arguments that the class of functions

$$\bigcup_{h>0} \left\{ x \mapsto \int_{\mathbb{R}} q(x-t-y) K_h^{(0)}(y)\, dy\colon\ |t| \geqslant \zeta \right\}$$

is in the $L^2(P)$-closure of $\|K\|_{L^1}$-times the symmetric convex hull of the P-pregaussian class $\bar{\mathcal{F}} = \{q(\bullet - t)\colon\ |t| \geqslant \zeta/2\}$. To see this one can either make a minor modification of the argument in Lemma 1 in [10], or notice that, $\{q(\bullet - t)\colon\ |t| \geqslant \zeta/2\}$ being bounded in the separable Banach space $L^2(P)$ (cf. after (4.18)), the integrals $\int q(\bullet - t - y) K_h^{(0)}(y)\, dy$ are $L^2(P)$-valued Bochner-integrals, and can thus be obtained as $L^2(P)$-limits of simple functions lying in the symmetric convex hull of $\{z \mapsto \|K\|_{L^1} q(z-t)\colon\ |t| \geqslant \zeta/2\}$ (e.g., Appendix E and Theorem E.3 in [8]).

(e) Write f, g for distinct translates of q (elements of $\mathcal{F}$), and deduce from Minkowski's inequality for integrals that

$$\begin{aligned}\left(E\left[\left(f * K_h^{(0)}(X) - g * K_h^{(0)}(X)\right)^2\right]\right)^{1/2} &\leqslant \int_{-\zeta/2}^{\zeta/2} \left|K_h(u)\right| \left\| f(-u - \bullet) - g(-u - \bullet)\right\|_{2,P}\, du \\ &\leqslant \|K\|_{L^1} \sup_{|u| \leqslant \zeta/2} \left\| f(u - \bullet) - g(u - \bullet)\right\|_{2,P}.\end{aligned}$$

Since entropy bounds are preserved under Lipschitz transformations, and since

$$\left\{q(u - \bullet - t)\colon\ |t| \geqslant \zeta,\ |u| \leqslant \zeta/2\right\} \subseteq \left\{q(u - \bullet - t)\colon\ |t| \geqslant \zeta/2\right\}$$

has polynomial $L^2(P)$-covering numbers by the same arguments as after (4.18), we deduce the bound $H(\tilde{\mathcal{F}}_n, L^2(P), \eta) \lesssim \log(\eta^{-1})$ for every $\eta > 0$ small enough, independent of n. Conclude that we can take $\lambda_n(\eta) = \log(\eta^{-1})\eta^2$, so that the envelope condition (4.17) becomes

$$h_n^{-\alpha'} \lesssim (\log n)^{-1/2} n^{1/4}, \tag{4.19}$$

which is satisfied due to $\alpha' < \alpha$ and $h_n^{-\alpha} \lesssim n^{1/4}$, completing the proof.

Acknowledgments

The authors would like to thank an anonymous referee, Jakob Söhl and Mathias Trabs for a careful reading of the manuscript and several helpful suggestions and corrections, as well as Chris Klaassen for useful discussions.

References

[1] H. Bauer, Probability Theory, de Gruyter, Berlin, 1996.
[2] D. Belomestny, Spectral estimation of the Lévy density in partially observed affine models, Stochastic Process. Appl. 121 (1) (2011) 1217–1244.
[3] D. Belomestny, M. Reiß, Spectral calibration for exponential Lévy models, Finance Stoch. 10 (2006) 449–474.
[4] B. Buchmann, R. Grübel, Decompounding: an estimation problem for Poisson random sums, Ann. Statist. 31 (2003) 1054–1074.
[5] F. Comte, V. Genon-Catalot, Nonparametric adaptive estimation for pure jump Lévy processes, Ann. Inst. H. Poincaré Probab. Statist. 46 (2010) 595–617.
[6] V.H. de la Peña, E. Giné, Decoupling. From Dependence to Independence, Springer, New York, 1999.
[7] R.M. Dudley, Sample functions of the Gaussian process, Ann. Probab. 1 (1973) 66–103.
[8] R.M. Dudley, Uniform Central Limit Theorems, Cambridge University Press, Cambridge, 1999.
[9] G.B. Folland, Introduction to Partial Differential Equations, second ed., Princeton University Press, Princeton, 1995.
[10] E. Giné, R. Nickl, Uniform central limit theorems for kernel density estimators, Probab. Theory Related Fields 141 (2008) 333–387.
[11] E. Giné, R. Nickl, Uniform limit theorems for wavelet density estimators, Ann. Probab. 37 (2009) 1605–1646.
[12] E. Giné, J. Zinn, Some limit theorems for empirical processes, Ann. Probab. 12 (1984) 929–989.
[13] M. Girardi, L. Weis, Operator-valued Fourier multiplier theorems on Besov spaces, Math. Nachr. 251 (2003) 34–51.
[14] S. Gugushvili, Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process, J. Nonparametr. Stat. 21 (3) (2009) 321–343.
[15] G. Jongbloed, F.H. van der Meulen, A.W. van der Vaart, Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes, Bernoulli 11 (2005) 759–791.
[16] Y. Katznelson, An Introduction to Harmonic Analysis, Dover, New York, 1976.
[17] C.A.J. Klaassen, E. Veerman, Private communication, 2011.
[18] K. Lounici, R. Nickl, Global uniform risk bounds for wavelet deconvolution estimators, Ann. Statist. 39 (2011) 201–231.
[19] M.H. Neumann, M. Reiß, Nonparametric estimation for Lévy processes from low-frequency observations, Bernoulli 15 (1) (2009) 223–248.
[20] R. Nickl, Empirical and Gaussian processes on Besov classes, in: E. Giné, V. Koltchinskii, W. Li, J. Zinn (Eds.), High Dimensional Probability IV, in: IMS Lecture Notes, vol. 51, 2006, pp. 185–195.
[21] R. Nickl, B.M. Pötscher, Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type, J. Theoret. Probab. 20 (2007) 177–199.
[22] D. Radulović, M. Wegkamp, Weak convergence of smoothed empirical processes. Beyond Donsker classes, in: E. Giné, D.M. Mason, J.A. Wellner (Eds.), High Dimensional Probability II, in: Progr. Probab., vol. 47, Birkhäuser, Boston, 2000, pp. 89–105.
[23] K.-I. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, Cambridge, 1999.
[24] J. Schmidt-Hieber, A. Munk, L. Dümbgen, Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features, 2012, preprint, arXiv:1107.1404.
[25] M. Trabs, Calibration of selfdecomposable Lévy models, preprint, arXiv:1111.1067, 2012.
[26] H. Triebel, Theory of Function Spaces, Birkhäuser, Basel, 2010 (reprint of the 1983 edition).
[27] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.